Artificial Intelligence and R&D Alliances: Innovation Complements

Bowen Lou
School of Business
University of Connecticut
bowen.lou@uconn.edu

Evan Rawley
School of Business
University of Connecticut
evan.rawley@uconn.edu

Abstract

Firms increasingly rely on strategic alliances and artificial intelligence to innovate in knowledge-intensive environments. In this paper, we examine whether these two innovation-enhancing activities are complements, as suggested by contract theory. Using a rich dataset on the biopharmaceutical industry from 2010 to 2019, we show that firms with greater AI resources form 7% more R&D alliances and enjoy an 24% increase in new drugs developed from their alliances. Moreover, we observe a key mechanism behind the results: AI is particularly useful in exploiting information held by counterparties within an alliance. The study sheds light on how modern organizational and technological advancements jointly shape the innovation production function.

VISIT…

LANZAROTE Caliente.com

Introduction

This paper examines how artificial intelligence (AI) and research and development (R&D) alliances jointly influence innovation. Scholars have long recognized that strategic alliances facilitate inter-firm knowledge flows (Mowery et al. 1996), particularly in R&D-intensive environments (Hagedoorn 2002; Sampson 2007). More recently, advances in AI promise to improve the efficiency of R&D by directing scientists toward promising pathways for innovation (Cockburn et al. 2018). But, while both AI and R&D alliances are thought to facilitate knowledge-driven approaches to innovation, little is known about the interplay between these activities. And, yet AI's predictive power seems particularly well suited to enhance knowledge transfers between allied firms. Thus, this paper develops and tests the idea that AI and R&D alliances are complements in the innovation process.

Empirically, we study the biopharmaceutical industry, which is characterized by widespread R&D alliances and the rapid adoption of AI (Aggarwal and Hsu 2009; Lou and Wu 2021). Within the industry, it is common for large integrated pharmaceutical firms to pursue R&D alliances to access to high-potential novel chemical entities (NCEs) identified by small, focused biotechnology firms. To help transform an NCE into a cure, pharmaceutical firms use machine learning—AI software that improves the predictive power of a statistical model by continuously incorporating new data—to match the NCE to potential disease targets (e.g., Fleming 2018). Thus, the biopharmaceutical industry appears to be a natural context for studying how R&D alliances and AI jointly shape the production of knowledge.

The main results show that firms with greater AI resources form 4% more R&D alliances and generate 8% more drugs per alliance. The baseline differences-in-differences results are robust to standard econometric techniques for adjusting for endogeneity bias. We also demonstrate

that firms with greater AI resources are better at exploiting information embedded in their alliance relationships, which suggests that improved knowledge flows are, indeed, a key mechanism behind the complementary relationship.

As the importance of artificial intelligence grows within firms, scholars have taken a keen interest in understanding how AI influences firm strategy (Furman and Seamans 2019). This paper contributes to the literature on AI and strategy, by explicating how AI and strategic alliances jointly influence the innovation process. Our key insight—that by facilitating productive knowledge flows within an alliance, AI resources make R&D alliances more efficient—also contributes to contract theory. Scholars have long considered how characteristics of transactions and capabilities influence the relative efficiency of discrete structural alternatives (e.g., Williamson 1991; Hitt 1999). We show that a firm's analytical capability, embedded in its AI resources, influences the efficiency of transactions (i.e., alliances) themselves. Furthermore, by documenting the AI resources and R&D alliance complementarity, we contribute to the empirical literature on complementarities in the organization of innovation (e.g., Cassiman and Veugelers 2006). Finally, by providing evidence that AI improves knowledge flows within alliances, we help unpack one mechanism for how innovative firms use AI to gain competitive advantage. In sum, the paper sheds new light on how modern organizational and technological advancements are jointly reshaping the innovation production function, with implications for scholars and practitioners alike.

Theory and Hypothesis

Alliances are often considered "hybrid" transactions, as they have characteristics of both market exchange and vertical integration (Williamson 1991). Specifically, alliances combine some of the high-powered incentives of markets with some of the monitoring capabilities and administrative controls of integration (Mowery et al. 1996). Alliances are not a panacea—incentive problems

can arise, and they can be cumbersome to manage. Still, given the well-known problems of innovating within bureaucracies (e.g., Schumpeter 1934), and arm's length exchange in markets for innovation (Arrow 1962), alliances have emerged as an important organizational strategy for facilitating innovation (Arora et al. 2001; Arora et al. 2002).

R&D alliances are of particular interest because of their potential to nurture and develop novel ideas. And, indeed, prior studies have demonstrated that alliances can improve a firm's innovativeness (e.g., Rosenkopf and Almeida 2003), particularly when well-managed interfirm linkages create opportunities for firms to access meaningful information from their partners (Powell 1998). Yet, the benefits of R&D alliances depend not only upon the nature of the relationship, but also the focal firm's own capacity for utilizing the information embedded in the relationship (Cohen and Levinthal 1990). Therefore, firms with limited capacity for exploiting knowledge held by alliance partners effectively face larger frictions in the market for innovation, and, as a result, should form fewer alliances R&D alliances, and be less productive in the alliances they do form, *ceteris paribus*.

On the other hand, AI offers a powerful potential complement to R&D alliances. The literature on information technology (IT) and contract theory, shows that IT tends to reduce coordination costs, leading to shifts in the boundary of the firm (Hitt 1999; Rawley and Simcoe 2013). Therefore, it seems reasonable to assume that AI software could also influence the relative efficiency of certain types of transactions (i.e., alliances). Indeed, we know that AI eases frictions in knowledge flows between firms, by improving predictive search of complex shared data (Agrawal et al. 2018; Wu et al. 2019). Thus, our theory focuses on the complementarity between R&D alliances and AI resources embedded in technology and human capital.

To summarize, our baseline assumptions, which we later verify, are that innovative output is an increasing function of (i) the number of R&D alliances a firm is involved in, and (ii) a firm's AI resources. Our key proposition, though, is that R&D alliances will be more productive when a firm has higher levels of AI resources, and, equivalently, that a firm's AI resources will be more productive when the firm has more strategic alliances to work with. Therefore, our central hypothesis is that AI resources and R&D alliances exhibit complementarities, in the sense of Milgrom and Roberts (1990) and Cassiman and Veugelers (2006), and summarized by equation (1):

(1) Innovation (R&D alliances, AI) > Innovation (R&D alliances) + Innovation (AI).

In words, we propose Hypothesis 1:

(H1): Firms are more innovative when they jointly increase their AI resources and R&D alliances.

Hypothesis 1 is a prediction about how AI resources improves the intensive margin of R&D alliances—it makes R&D alliances more productive. A corollary to the theory considers the extensive margin: firms with greater AI resources should form more R&D alliances too. The corollary follows directly from the idea that firms will rationally exploit the intensive margin complementarity between AI and R&D alliances by expanding on the extensive margin, creating more potential opportunities where the analytical power of AI can be utilized. Thus, Hypothesis 2 predicts:

(H2): Higher levels of AI resources leads to more R&D alliances.

Hypothesis 2 follows directly from the logic for Hypothesis 1—we know that if two corporate practices are complements, those practices should tend to cluster (Arora and Gambardella 1990; Brynjolfsson and Milgrom 2013)—but the empirical test of Hypothesis 2 also challenges the theory in a different and important way. "True" complementarity implies that AI resources improve R&D alliances, and that R&D alliances improve the usefulness of AI resources. But, finding that firms with greater AI resources have more productive R&D alliances—a finding in support of Hypothesis 1—would also be consistent with a scenario where AI resources help firms screen potential R&D alliances more effectively. While it might also be interesting to consider how AI resources helps firm screen out low performing potential alliance partners, our theory is about increasing innovation not filtering. Evidence in support of Hypothesis 2 helps distinguish between our theory and a plausible alternative explanation.

Finally, we also make explicit our hypothesis about the mechanism underlying the complementarity between AI resources and R&D alliances. Our theory predicts that AI resources complements R&D alliances because it enables the focal firm to better exploit knowledge embedded in its alliance partner. Therefore, if our theory is correct, we should be able to detect greater knowledge flows within alliances when focal firms have greater AI resources. Hypothesis 3 summarizes the prediction:

(H3) Increasing AI resources improves knowledge flows within an R&D alliance.

The three hypotheses, proposed above, form a testable set of propositions that help triangulate a simple theory of complementarity. R&D alliances provide firms with potential knowledge, but exploiting that knowledge is non-trivial. AI is a useful analytical tool, but its effectiveness is limited by the extent of data available to the user of the AI. By facilitating knowledge sharing across firm boundaries, AI resources and R&D alliances complement one another, creating real opportunities for meaningful innovation.

Institutional Context

The biopharmaceutical value chain consists of several phases. The first is basic research where novel drug entities are tested against certain biological targets in a laboratory setting. The next phases are stages of drug development. The first step in development is non-clinical (testing on non-human animals). The next three phases involve human clinical trials of increasingly larger sizes. Finally, a successful drug gains FDA approval, is marketed and sold. Our study focuses on the drug discovery through the non-clinical development stage, as that is where the impact of AI and R&D alliances on industry innovation are thought to be the greatest (e.g., Savage 2021).

The industry is an excellent laboratory to test our theory for two main reasons. First, the industry relies on a wide range of knowledge-intensive interorganizational activities to foster innovation. Over the past decades, biopharmaceutical firms have entered into systematic collaborative linkages with universities, and other research-intensive firms, which provide complementary assets necessary for innovation (Arora and Gambardella 1994). As alliances have proven to be a potential source of competitive advantage, the scope and scale of interorganizational collaborations and alliances in this sector have grown (Powell et al. 1996). As a result, successful alliances typically require extensive knowledge transfers between firms (Oxley and Sampson 2004;

Sampson 2007). Therefore, the biopharmaceutical industry is a good laboratory for studying the linkages amongst knowledge management, strategic alliances and innovation.

Second, as knowledge needed for discovering drug candidates is becoming more codifiable, through the process of digitization, the importance of AI has grown within the industry. AI is most effective with large structured sources (Agrawal et al. 2018), where it can be used to find subtle linkages amongst data elements to make useful predictions about potential innovation pathways (Wu et al. 2020). Hence, the potential drug candidates screened using AI serve as a useful starting point to improve efficiency in the drug innovation process. For example, machine learning has been applied to produce the knowledge graph for hypothesis generation and validation, and to infer the interplay between numerous biological entities such as genes, proteins, diseases and drug candidate, which may fundamentally reshape the innovation production function in drug discovery (Fleming 2018).

In sum, we study a context where innovation and knowledge management are crucial, and strategic alliances, and AI are known to independently influence innovation. More importantly for the tests of our theory, it is also plausible that AI and strategic alliances might complement one another in the biopharmaceutical industry. To that end, we focus on the interplay between R&D alliances and AI in the innovation production function.

Data and measures

We study the biopharmaceutical industry from 2010 to 2019, a period that witnessed rapid growth in the application of AI, as well as the proliferation of alliance activities. Our main dependent variable is *Drugs developed*, measured as the cumulative number of drug candidates developed by a firm through the pre-clinical phase, depreciated at 15% per year, as is common practice in the innovation literature (Hall 1990). At the pre-clinical phase a drug has proven scientific potential,

and is a serious contender to become a product. Thus, developing a drug to the pre-clincal phase represents a meaningful innovative step. Although many drugs fail during clinical trials, there can be no clinical trials until a drug reaches the pre-clinical phase.

Data on drugs developed are collected from two sources that have been used extensively to study biopharmaceutical innovation (e.g., Sosa 2014 and Krieger et al. 2022): the Informa Pharmaprojects database, and the investigational drug database from Clarivate Analytics databases. The data are very rich, but they do have one important limitation: we cannot map alliances to specific drugs. Therefore, we study changes in a firm's stock of drugs developed to capture the average complementarity effect.

The data do map drugs to owners, however, because ownership of a drug can be changed through mergers and acquisitions (M&A), we use SDC Platinum and Zephyr data from the Bureau van Dijk to properly match drugs to the firms responsible for their development. After cleaning the ownership field, we have 614 unique firms that developed at least one drug through the pre-clinical phase between 2010 and 2019.

Data on *R&D Alliances* in the biopharmaceutical industry, our first key explanatory variable, are collected from the Biosci database (BiosciDB). The database tracks interfirm R&D contracts, providing the most comprehensive records of biopharma R&D alliances since the early 1980s.[1] There are 2,755 R&D alliances in the data. We cumulate R&D alliances, at the firm level, to capture a firm's overall exposure to R&D alliances.

Our second key explanatory variable is *Artificial Intelligence (AI) resources*, which is measured as a function of the count of a firm's AI-related patents (*AI IP*) and a count of a firm's

[1] Compared to some alternative alliance databases such as Securities Data Company (SDC) Platinum database on Joint Ventures and Alliances, the BiosciDB database offers wider coverage of deals in the biotechnology and pharmaceutical sector.

AI-related workers (*AI skills*). *AI IP* is an intangible asset-based measure of AI resources (Webb 2019). To measure a firm's AI IP, we use global patents from a worldwide patent statistical database, PATSTAT, created by the European Patent Office (EPO).[2] PATSTAT offers bibliographical data for over 100 million patents from 90 patent-issuing authorities from both leading industrialized and developing countries going back as far as the nineteenth century. It provides patent records with rich information that contains patent application year, patent technology classes, as defined by the International Patent Classification (IPC) or Cooperative Patent Classification (CPC) systems, citations, title, abstract, and legal entities (e.g., firms or any organizations) filing the patent application. Patent owners are identified using a standardization of the original name of the patentees. We use the patent filing year, as opposed to the publication year, because it more closely approximates the time when the firm produced the innovation described in the patent (Griliches et al. 1986; Hall et al. 2005).

To create a robust measure of a firm's *AI IP*, we take four steps to identify whether a patent uses or contributes to AI technologies. We start with the most obvious AI patents: USPTO class 706 for "Data Processing – Artificial Intelligence" consists of a large set of subclasses including "machine learning" and "neural network". To link USPTO class 706 to the PATSTAT data, we apply the IPC or CPC concordance to obtain the classification code that is used for classifying global AI patents.

While class 706 explicitly captures patents at advance the knowledge frontier of AI, the narrow scope of the USPTO AI classification code may exclude patents employing AI to solve

[2] We leverage the PATSTAT data in the version 2017b to cover patent records up to 2017. As Google launched its public datasets of worldwide patents on BigQuery in 2017 (https://cloud.google.com/blog/products/gcp/google-patents-public-datasets-connecting-public-paid-and-private-patent-data), we further augment our patent data to cover the recent time period from 2017 to 2019 by using global patents through Google Patents Public Datasets. We adopt similar approaches to match them to other datasets used in our empirical study.

problems in an innovative way, which do not contribute to fundamental advances in AI. Thus, our second step expands the scope of AI patents to include content-based information available in the patent documents. Specifically, we follow Cockburn et al. (2018), who define three interrelated technological subfields within AI to characterize the evolution of achievements in AI: robotics, symbolic systems, and learning. Using their classification scheme, we construct a comprehensive list of validated words and phrases pertaining to AI and search for these terms in both the titles and abstracts of our patents. (See the Appendix for keywords used).

Our third step is to include patents with content that fall within the well-accepted Association of Computing Machinery (ACM) Computing Classification System (CCS), which accounts for dynamic changes in AI technologies over time (WIPO 2019). The ACM/CCS methodology has been used for over 50 years to organize technological concepts and trends. Specifically, we use three major hierarchies in CCS that identify AI-related phrases, including: (i) the "artificial intelligence" hierarchy, comprised of an AI functional application, such as natural language processing, computer vision, knowledge representation and reasoning, simulation of human cognitive tasks, and AI techniques used to realize those functions; (ii) a "machine learning" hierarchy, which includes numerous learning-based AI techniques; and (iii) a "life and medical sciences" hierarchy under the "applied computing" category that captures activities concerning intelligent computing for producing medicines.

Our fourth step is to expand the vocabulary list from ACM/CCS to contain phrases related to a variety of AI technologies specifically used for drug innovation in more recent years. Specifically, we include well-known off-the-shelf AI tools and systems such as PyTorch and TensorFlow (Raymond et al. 2019).

Based on the four steps described above, we track the development of *AI IP* over time, as the count of all AI patents, depreciating the value at 15% per year. The resulting AI stock variable aligns with the spirit of classic innovation production functions that model innovation as a function of the existing repertoire of knowledge and research resources dedicated to the production of new knowledge (Romer 1990).

Demand for personnel with AI skills represents a human capital-based measure of AI resources, as it captures the personnel-based resources firms seek to create, implement, and deploy in AI-related areas within the firm. To measure *AI skills*, we use job titles and skill requirements data from a company that collects the data from over 40,000 online job boards worldwide (Alekseeva et al. 2021). Following our systematic approach to classifying AI patents, we search for AI-related words in each job posting for each firm in our sample. Additionally, job title classifications from O*NET are also applied to identify AI-related positions, similar to the way information technology and analytics labor has been identified in prior research studies (e.g., Tambe and Hitt (2012)). Once individual AI job postings are identified, we aggregate *AI skills* to the firm level as a cumulative count.

Although *AI IP* and *AI skills* are independent measures of a firm's observable AI resources, they are almost surely jointly responsible for a firm's AI resources. For example, AI IP may proxy for whether the firm has: (1) tangible AI resources, such as the technological infrastructure required to develop an AI-related patent, or (2) intangible AI resources, such as the practices required to use AI for innovation (Lou and Wu 2021), or both. And, a firm's AI human capital, proxied by AI skills, are almost surely involved with the development of both tangible and intangible AI resources. Thus, we allow *AI IP* and *AI skills* (and their interaction with *R&D Alliances*) to enter the regressions separately in some specifications, but aggregate them into a

single composite AI resources measure for our main tests, by standardizing and summing each AI-related dimension, as shown in equation (2):

$$(2)\ AI\ resources = norm(norm(AI\ IP) + norm(AI\ skills)).$$

The result is a normalized measure of a firm's AI resources that jointly considers both its asset-based and human capital-based AI resources. There are 185 firms in our sample with positive AI resources in at least one year.

Alliances serve as conduits through which information flows to a focal firm from its partner. To better understand the mechanism underlying the proposed innovation complementarity between AI and R&D alliances, we focus on the 2,755 alliances in our data set, using the alternative dependent variable *Knowledge flows*, which draws on citation patterns in an alliance firm's patent portfolio to proxy for the changes in the relationship of its knowledge portfolio to that of its allied firm (Mowery et al. 1996). The test examines the extent to which knowledge can be leveraged by client firms (typically vertically integrated traditional pharmaceutical or large "platform" biotech companies) via their R&D partners (typically small biotechnology companies). Following recent work on patents and alliances, we use patent citation data to capture the changes in the citation patterns of focal firms before and after they form alliances to trace knowledge flows from its partners (Almeida et al. 2002; Oxley and Wada 2009).

The alliance data classifies firms as "R&D" companies, firms that are primarily responsible for basic research, or "client" companies, firms that are accessing knowledge from their collaborators in the pursuit of drug development (Aggarwal and Hsu 2009; Lee et al. 2015). This dichotomy has been well defined in both academic and practical analysis of the R&D partnerships

in the biotechnology and pharmaceutical sector (e.g., Robinson and Stuart 2007a; Robinson and Stuart 2007b). Thus, for each of the R&D alliances in our sample, we calculate knowledge flow as the log difference between a focal client's pre- and post-alliance citations to its partner R&D firm, for the two years before and after the alliance is consummated.[3] We define the resulting measure as *Knowledge flows* to the client firm. The variable has a mean of 0.44, a range of zero to 678, and a standard deviation of 13.65. If our theory is correct, we should find that firms with greater AI resources should enjoy greater knowledge flows from their alliance partners.

Control Variables: Three databases—Crunchbase, PitchBook and Bureau van Dijk Orbis—are used to retrieve other firm characteristics such as firm ownership status, firm age, number of employees and R&D expenses. The databases provide a wide coverage of both public and private bio-pharmaceutical firms. We account for a firm's financial ownership status over years (variable *Public Status* = 1 when the firm is publicly held, otherwise *Public Status* = 0), which has been shown to influence organizational innovation outcomes (Aggarwal and Hsu 2013). Additionally, we control for firm age, which is the difference between the observation year and the founding year of each firm in our sample. The oldest firm, Merck, is 347 years old, but the mean age is 23.81 years old, as there are many young biotechnology companies in the sample as well. We also account for each firm's size (by number of employees) and R&D expenditures, as well as the count of a firm's overall number of patents, depreciated at 15% per year.

Empirical Strategy

In the ideal experiment we would randomly assign AI resources and R&D alliances and evaluate the effect of jointly adopting different levels of each on innovation. However, our analysis must instead deal with a non-random data generating process where firms choose both their level of AI

[3] Using other year lags such as one-year and three-year lags produces similar results.

resources and R&D alliances. The endogeneous selection by firms into AI and R&D alliances represents a major challenge to a causal interpretation of the results. For example, if better (worse) firms devote more resources to AI and R&D alliances, "naïve" estimates of their joint effect on the innovation production function will be upward (downward) biased.

To address the most fundamental endogeneity concerns, we deploy a differences-in-differences estimator, using firm and time fixed effects, that controls for all time invariant heterogeneity across firms, including firm quality differences, as well as any secular time trends that influences all firms. Firm fixed effects also eliminate any differences in initial stocks of drugs developed, AI resources, R&D alliances, allowing for an apples-to-apples comparison of the effect of *changes* in AI resources and R&D alliances on *changes* in drugs developed. As a result, our baseline analyses can be interpreted as treatment on the treated (ToT) estimators, revealing whether AI resources and R&D alliances are complements for firms that choose to use them together.

While ToT effects, from differences-in-differences estimators, are interesting, ideally, we would like to know the average treatment effects (ATE) of jointly increasing AI resources and R&D alliances. If firms choose to jointly increase AI resources and R&D alliances, based on unobservable (to the econometrician) time-varying firm-specific characteristics that are correlated with both innovation and the interaction between AI resources and R&D alliances, then ToT effects will diverge from ATE. For example, one could reasonably be concerned that firms might adopt AI and R&D alliances when they expect to have more new drug candidates in the future, biasing our ToT estimates upward, compared to "true" ATE. Alternatively, in anticipation of an innovation drought in the future, firms might ramp up their AI resources and R&D alliances in the

hopes of reenergizing their discovery process, biasing the ToT estimates downward, compared to ATE.

To deal with the more subtle endogeneity inherent in time-varying firm-specific covariation between innovation and our key explanatory variables, we harness two standard statistical corrections: propensity score matching and a two-stage least squared (2SLS) instrumental variables analysis. Our 2SLS approach uses the total number of neighboring firms, defined as firms that cite each other's patents, that have AI resources, as an instrument for the focal firm's AI resources. The rationale for the instrument is that if many firms, working in the same scientific area, feel the need for AI resources to make progress, then the focal firm is partially induced to increase their AI resources, relative to firms working in scientific areas where AI is less common. If competition essentially forces a focal firm to increase their level of AI resources, a portion of the change in AI resources ceases to be a choice variable, allowing us to estimate the causal effect of AI resources using a 2SLS estimator.

To be effective and valid an instrument should be strongly correlated with the endogenous regressor (i.e., AI resources), and should solve the exclusion restriction—it should not impact a focal firm's outcomes directly. Although using similar firms' characteristics as instruments is a common econometric technique (e.g., Nevo 2001; Campa and Kedia 2002), there is no way to formally test the exclusion restriction. Still, unless rival firms' AI resources degrades a focal firm's innovativeness, by outpacing them in the race for specific innovations—an assumption we do test directly below—it seems reasonable to assume that other firm's AI decisions would not directly influence the innovativeness of a focal firm.

Unfortunately, we could not identify an instrument for R&D alliances that both met the exclusion restriction and was strong in the first stage. Therefore, we rely on a propensity score

matching estimator to deal with endogenous selection into different levels of alliance formation. Propensity score estimators match firms based on their pre-treatment characteristics to ensure that one does not conflate treatment effects with observable heterogeneous pre-trends. In the absence of systematic unobservable heterogeneity that is correlated with both the treatment (i.e., R&D alliances) and the outcome of interest, propensity score estimators can be considered "as good as" average treatment effects (Rosenbaum and Rubin 1983).[4]

Main specifications

Our baseline differences-in-differences estimator captures the relationship between changes in R&D alliances, in conjunction with changes in AI resources, and changes in innovation. Thus, we want to match firms that increase their number of alliances to firms that do not, contemporaneously (i.e., in any given year). To implement our matching estimator, we use a first stage probit predicting whether a firm forms an alliance in a particular year, based on its *ex ante* characteristics, and use the predicted value from the probit—the probability of being treated, or the "propensity score"—to match control and treatment groups firms, one to one, using nearest neighbor matching without replacement. The resulting matched sample allows us to compare firms that are equivalent on all observable characteristics, at any given point in time, except for their decision to increase the number of alliances in that year, or not. While propensity score matching cannot control for unobservable differences between firms, given our baseline differences-in-differences estimation strategy, it would appear to provide a reasonable basis for inferring average treatment effects.

We begin our empirical analysis by examining the potential innovation complementary between AI and R&D alliances, at the firm-year level, for firm *i* in year *t*, with the differences-in-differences estimator in specification (3):

[4] Other matching and weighting approaches, such as coarsened exact matching (CEM) and inverse probability weighting (IPW), yield consistent results.

$$(3)\ Drugs\ developed_{it} = \beta_0 + \beta_1 AI\ resources_{it} + \beta_2 Alliances_{it} + \beta_3 AI\ resources_{it} * Alliances_{it} + X_c\beta_c + T_t + \gamma_i + \varepsilon_{it},$$

which includes firm fixed effects γ_i, and year fixed effects T_t. Subscript c indexes the vector of coefficients on time-varying firm-specific controls, including patent stock, ownership status, firm age, workforce size and R&D spending. Standard errors are clustered by firm. The coefficient β_1 represents the main effect of increasing a firm's AI (IP, skills, or aggregate resources), β_2 captures the main effect of increasing the number of R&D alliances, and β_3, the coefficient on the interaction term between AI and R&D alliances, is the key coefficient of interest.

Hypothesis 1 predicts an innovation complementarity between AI resources and R&D alliances, on the intensive margin—each alliance is more productive when consummated by a firm with greater AI resources. Hypothesis 2 predicts we should also expect to see an extensive margin complementarity—firms with greater AI resources should pursue more R&D alliances, in the expectation that each alliance should be more innovative. To test whether there is indeed an extensive margin complementarity, we examine the conditional correlation between AI resources and R&D alliances. Since AI raises the benefit of information obtained from any given alliance, there should be a positive correlation between AI resources and number of R&D alliances. Thus, we expect $\beta_1 > 0$ in specification (4).

$$(4)\ Alliances_{it} = \beta_0 + \beta_1 AI\ resources_{it} + X_c\beta_c + T_t + \gamma_i + \varepsilon_{it}.$$

Our theory predicts a complementarity between AI resources and R&D alliances because R&D alliances provide firms with new data, while AI resources allow firms to utilize advanced software systems to improve data analysis, facilitating the creation of new knowledge. To test if the knowledge creation mechanism underlies the innovation result, as predicted by the theory, we examine interfirm knowledge flows at the transaction (i.e., alliance) level.

For each alliance formed by a focal "client" pharmaceutical firm *i*, beginning in year *t*, we regress the knowledge flows from its R&D partner on its AI resources, controlling for *i*'s other characteristics, as described in the firm-level analysis. Specifically, the mechanism test specification is (5):

$$(5)\ \textit{Knowledge flows}_{it} = \beta_0 + \beta_1 \textit{AI resources}_{it} + X_c\beta_c + T_t + \gamma_i + \varepsilon_{it}.$$

If $\beta_1 > 0$, it would suggest that AI resources facilitates knowledge flows from R&D companies (typically biotechnology companies) to client pharmaceutical companies.

Results

We begin by visualizing a preview of our main finding in Figure 1. The graphic reveals that firms involved in above the median level of R&D alliances are more innovative, but the most innovative firms also have high levels of AI resources. The figure provides model-free evidence consistent with the idea that R&D alliances and AI resources are complements. Whether the relationship is casual or not is the subject of much of the ensuing analysis.

Table 1 displays the summary statistics and bivariate correlations amongst the variables. The mean number of drugs developed is 15.33 per firm-year. The mean values of *AI IP* and *AI skills* are 1.22 and 16.12, respectively, while the mean number of alliances is 0.61. As one might

expect, firms spend enormous sums on R&D—the mean value is $3.56 billion per firm-year. In the regressions, we take the natural logarithm of a firm's number of drugs, alliances, AI IP and skills, and all the non-categorical controls to account for their skewness, and to facilitate interpretation of the results as elasticities.

In Table 2, we study the impact of AI and R&D alliances on innovation performance. For presentation purposes, all coefficients and standard errors are multiplied by 100, in Table 2 and all subsequent tables, so one can interpret the coefficient estimates as elasticities in percentages. We find that the main effects of a firm's AI IP and number of R&D alliances are positively correlated with the number of drugs developed (Column 1 in Table 2). Including the interaction of AI IP with R&D alliances, does not change the sign or significance of the main effects, while the coefficient on the interaction term is positive and significant (Column 2 in Table 2). The preliminary interpretation, subject to further econometric tests, is that AI resources embodied in a firms AI-related patents is an effective tool for exploiting external knowledge. Replacing AI IP with AI skill, and running parallel tests, produces similar results (Columns 3 and 4, Table 2). The preliminary interpretation is that workers with AI skills are better at exploiting external knowledge to improve innovation performance.

Next, we combine the AI IP and AI skill measures into a compositive index of AI resources and run another set of parallel results. The sign, significance, and approximate magnitude of the results are the same (Columns 5 and 6, Table 2). The elasticity of the main effect of R&D alliances is 6.73%, the elasticity on AI resources is 6.77% and the elasticity of their interaction is 4.17%. The (preliminary) economic interpretation is that doubling either AI resources or R&D alliances independently, is only about 76% as effective as doubling them together.[5] Taken together with

[5] Doubling the number of R&D alliances independently generates 6.73% more drugs, while doubling AI resources independently generates 6.77% more drugs. The sum of the two independent effects is 13.50%. Doubling R&D

the AI IP and AI skill results, the pattern of evidence suggests that there are complementarities between a firm's AI resources its R&D alliances.

To address concerns that R&D alliances are endogenous, we match firms with a new alliance to firms with no new alliances in year *t*, on all the observable, lagged one-year, *ex ante* characteristics of the firms available to us, using propensity score matching. Table 3 reports statistics of the observable characteristics before and after the matching. Before matching, the two groups are statistically different along every dimension. After matching, the two groups of firms have similar distributions, suggesting that selection bias is meaningfully reduced. Figure 2 shows the convergence of the propensity scores distributions graphically. We use the matched sample as the basis for our 2SLS "treatment effect" regressions.

Table 4 tabulates the 2SLS estimator, instrumenting for AI resources, using the number of neighboring firms with AI resources, within the matched R&D alliances sample. Since there are two endogenous variables remaining (i.e., after matching), *AI resources* and *AI resources x R&D alliances*, we include the main effect of the instrument and the instrument interacted with (the matched pairs of) R&D alliances. Thus, there are two first stage results to consider. The first set, (Column 1-I of Table 4) shows that neighboring firms' AI resources is a strong predictor of a focal firm's AI resources. Since our main concern with the validity of the exclusion restriction is that other firm's AI resources would *reduce* the focal firm's innovativeness (i.e., by allowing a rival to beat them in the race to develop a new potential drug), the first stage result appears to support the validity of the instrument.

The second set of first stage results, (Column 1-II of Table 4) reveals that the interacted instrument is also a strong predictor of the endogenous interaction term *AI resources x R&D*

---

alliances while doubling AI resources simultaneously, generates 4.17% additional drugs. 4.17% + 13.50%=17.67%. 13.50%/17.67% = 76%.

alliances. While the negative coefficient on the main effect of the instrument in Column 1-II is indicative of collinearity between the two instruments, the F-statistic on the joint significance of the instruments from the first stage is 27.85, suggesting that the instruments are jointly strong enough to generate meaningful statistical power in the second stage.

The second stage results in Column 2 of Table 4, shows that the point estimate of the 2SLS estimator is 7.51%, which is almost twice as large as the differences-in-differences estimator (see Column 6 of Table 2), though noisier. The interpretation is that there is a causal complementarity between AI resources and R&D alliances.

By analyzing the marginal effect of the AI resources and R&D alliances complementarity, the first set of analyses in Tables 2 and 4 focused on the intensive innovation margin. In Table 5, we study the extensive innovation margin, whether AI resources influence the number of R&D alliances. We do so with three sets of analyses. First, we regress the categorical variable *Alliance in year t*, which is set equal to one when an alliance is formed in a particular year *t*, and zero otherwise, on AI resources, with and without firm fixed effects. The estimation results suggest that firms with greater AI resources are more likely to enter an alliance in any given year (Column 1 of Table 5), even after adjusting for firm-specific heterogeneity (Column 2 of Table 5). Limited dependent variable models, probit and logit regressions, yield consistent results.

Second, we estimate a differences-in-differences estimator with the cumulative number of alliances as the dependent variable (Column 3 of Table 5). On average, a one-standard-deviation increase in a firm's AI resources is associated with a, precisely estimated, 6.68% increase in number of alliances. Although the evidence "merely" demonstrates a treatment on the treated effect on the extensive margin—firms expand the number of R&D alliances after increasing their

AI resources—it still provides support for the idea that R&D alliances and AI resources are complements.

Third, in columns 4-6 of Table 5, we show the 2SLS estimators for each of the specifications in columns 1-3, using the same instrument as in Table 4 column 1. The 2SLS point estimates are substantially larger and noisier than the OLS/fixed effects estimates, but the collective evidence points toward a causal effect of AI resources on R&D alliance formation.

Finally, to probe the mechanism through which the complementarity, we study knowledge flows around the time each R&D alliance is formed. Separate estimation of the two components of AI resources—AI IP and AI skills—yield positive associations (Columns 1 and 2 in Table 6). Using the composite measure of AI resources in Column 3 of Table 6, we find that, on average, a one-standard-deviation increase in AI resources is associated with about an 8% increase in its knowledge inflows, as captured by the difference in post-alliance and pre-alliance patent citations to its partners.

Column 4 of Table 6 shows the 2SLS estimator. The point estimate is about twice the differences-in-differences estimator, but not statistically different, due to the relatively large standard error. The interpretation is that AI resources cause knowledge flows to increase within R&D alliances, as predicted by the theory.

Conclusion

This research provides the conceptual and empirical basis for identifying a complementarity between a firm's AI resources and its R&D alliances in the firm's innovation production function. The main conceptual thrust is that by delivering a more capable analytical engine, AI resources reduce frictions inherent in markets for innovation (i.e., R&D alliances), allowing firms to capitalize on both the extensive margin, by making more use of the market for innovation (i.e.,

forming more R&D alliances), and on the intensive margin, by wringing more innovation out of each transaction (i.e., alliance).

The results show that doubling AI resources and R&D alliances in concert generates a 24% improvement in the innovation rate, compared to doubling AI resources and R&D alliances independently (the intensive margin effect); while doubling AI resources also leads to approximately 7% more R&D alliances (the extensive margin effect). Furthermore, we show that higher levels of AI resources are associated with more knowledge sharing, precisely the mechanism at the heart of our theory.

Taken together, the evidence suggests that R&D alliances allow firms with greater AI resources to better exploit alliance partner's knowledge to improve their innovativeness. While we believe this is the first large sample study to show the complementarity between AI an R&D alliances, we are certainly not the first to evaluate the relationship amongst IT, alliances and innovation. However, compared to prior work, this study offers both more conceptual and empirical precision, and more robust measures of resources. Prior work has treated IT as the amalgam of disparate communication technologies, computer software and hardware, whereas we focus explicitly on AI as an analytical tool. Moreover, our index of AI resources includes both asset- and human capital-based resources, whereas prior work has typically considered only one or the other. We also build on exemplar prior work, by taking identification seriously and by providing evidence of the mechanism at play.

Our findings have implications for market for scholars and practitioners alike. The search for R&D complementarities has been particularly fertile ground for managerially relevant research. This paper shows the importance of including AI resources and R&D alliance complementarities in that conversation.

References

Aggarwal, V. A., and Hsu, D. H. 2009. "Modes of Cooperative R&D Commercialization by Start-Ups," *Strategic Management Journal* (30:8), pp. 835-864.

Aggarwal, V. A., and Hsu, D. H. 2013. "Entrepreneurial Exits and Innovation," *Management Science* (60:4), pp. 867-887.

Agrawal, A., Gans, J., and Goldfarb, A. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press.

Alekseeva, L., Azar, J., Gine, M., Samila, S., and Taska, B. 2021. "The Demand for AI Skills in the Labor Market," *Labour Economics* (71), p. 102002.

Almeida, P., Song, J., and Grant, R. M. 2002. "Are Firms Superior to Alliances and Markets? An Empirical Test of Cross-Border Knowledge Building," *Organization Science* (13:2), pp. 147-161.

Arora, A., Fosfuri, A., and Gambardella, A. 2001. "Markets for Technology and Their Implications for Corporate Strategy," *Industrial and Corporate Change* (10:2), pp. 419-451.

Arora, A., Fosfuri, A., and Gambardella, A. 2002. "Markets for Technology in the Knowledge Economy," *International Social Science Journal* (54:171), pp. 115-128.

Arora, A., and Gambardella, A. 1990. "Complementarity and External Linkages: The Strategies of the Large Firms in Biotechnology," *The Journal of Industrial Economics*, pp. 361-379.

Arora, A., and Gambardella, A. 1994. "Evaluating Technological Information and Utilizing It: Scientific Knowledge, Technological Capability, and External Linkages in Biotechnology," *Journal of Economic Behavior & Organization* (24:1), pp. 91-114.

Arrow, K. 1962. "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, pp. 609-626.

Brynjolfsson, E., and Milgrom, P. 2013. "Complementarity in Organizations," *The Handbook of Organizational Economics*, pp. 11-55.

Campa, J. M., and Kedia, S. 2002. "Explaining the Diversification Discount," *The Journal of Finance* (57:4), pp. 1731-1762.

Cassiman, B., and Veugelers, R. 2006. "In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition," *Management Science* (52:1), pp. 68-82.

Cockburn, I. M., Henderson, R., and Stern, S. 2018. "The Impact of Artificial Intelligence on Innovation," 0898-2937, *National Bureau of Economic Research*.

Cohen, W. M., and Levinthal, D. A. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, pp. 128-152.

Fleming, N. 2018. "How Artificial Intelligence Is Changing Drug Discovery," *Nature* (557:7707), p. S55.

Furman, J., and Seamans, R. 2019. "AI and the Economy," *Innovation Policy and the Economy* (19:1), pp. 161-191.

Griliches, Z., Pakes, A., and Hall, B. H. 1986. "The Value of Patents as Indicators of Inventive Activity." *National Bureau of Economic Research* Cambridge, Mass., USA.

Hagedoorn, J. 2002. "Inter-Firm R&D Partnerships: An Overview of Major Trends and Patterns since 1960," *Research Policy* (31:4), pp. 477-492.

Hall, B. H. 1990. "The Manufacturing Sector Master File: 1959-1987," *National Bureau of Economic Research*.
Hall, B. H., Jaffe, A., and Trajtenberg, M. 2005. "Market Value and Patent Citations," *RAND Journal of Economics*, pp. 16-38.
Hitt, L. M. 1999. "Information Technology and Firm Boundaries: Evidence from Panel Data," *Information Systems Research* (10:2), pp. 134-149.
Krieger, J., Li, D., & Papanikolaou, D. (2022). "Missing novelty in drug development, " The *Review of Financial Studies*, 35(2), 636-679.
Lee, J., Hoetker, G., and Qualls, W. 2015. "Alliance Experience and Governance Flexibility," *Organization Science* (26:5), pp. 1536-1551.
Lou, B., and Wu, L. 2021. "AI on Drugs: Can Artificial Intelligence Accelerate Drug Development? Evidence from a Large-Scale Examination of Bio-Pharma Firms," *MIS Quarterly* (45:3).
Milgrom, J., and Roberts, J. 1990. "The Economics of Modern Manufacturing: Technology, Strategy and Organization." *The American Economic Review* 80(3): 511-528.
Mowery, D. C., Oxley, J. E., and Silverman, B. S. 1996. "Strategic Alliances and Interfirm Knowledge Transfer," *Strategic Management Journal* (17:S2), pp. 77-91.
Nevo, A. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica* (69:2), pp. 307-342.
Oxley, J., and Wada, T. 2009. "Alliance Structure and the Scope of Knowledge Transfer: Evidence from Us-Japan Agreements," *Management Science* (55:4), pp. 635-649.
Oxley, J. E., and Sampson, R. C. 2004. "The Scope and Governance of International R&D Alliances," *Strategic Management Journal* (25:8-9), pp. 723-749.
Powell, W. W. 1998. "Learning from Collaboration: Knowledge and Networks in the Biotechnology and Pharmaceutical Industries," *California Management Review* (40:3), pp. 228-240.
Powell, W. W., Koput, K. W., and Smith-Doerr, L. 1996. "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology," *Administrative Science Quarterly*, pp. 116-145.
Rawley, E., and Simcoe, T. 2013. "Information Technology, Producitivity and Asset Ownership: Evidence from Taxicabl Fleets." *Organization Science* 24(3): 831-845.
Raymond, P., Yoav, S., Erik, B., Jack, C., John, E., Barbara, G., Terah, L., James, M., Saurabh, M., and Carlos, N. J. 2019. "The AI Index 2019 Annual Report," AI Index Steering Committee, Human-Centered AI Institute, Stanford University.
Robinson, D. T., and Stuart, T. E. 2007a. "Financial Contracting in Biotech Strategic Alliances," *The Journal of Law and Economics* (50:3), pp. 559-596.
Robinson, D. T., and Stuart, T. E. 2007b. "Network Effects in the Governance of Strategic Alliances," *The Journal of Law, Economics, & Organization* (23:1), pp. 242-273.
Romer, P. M. 1990. "Endogenous Technological Change," *Journal of Political Economy* (98:5, Part 2), pp. S71-S102.
Rosenbaum, P. R., and Rubin, D. B. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* (70:1), pp. 41-55.
Rosenkopf, L., and Almeida, P. 2003. "Overcoming Local Search through Alliances and Mobility," *Management Science* (49:6), pp. 751-766.

Sampson, R. C. 2007. "R&D Alliances and Firm Performance: The Impact of Technological Diversity and Alliance Organization on Innovation," *Academy of Management Journal* (50:2), pp. 364-386.

Savage, N. "Tapping into the Drug Discovery Potential of AI." *Nature.com*. https://www.nature.com/articles/d43747-021-00045-7

Schumpeter, J.A. 1934. *The Theory of Economic Development*. Oxford University Press: London.

Sosa, M. L. 2014. "Corporate Structure, Indirect Bankruptcy Costs, and the Advantage of De Novo Firms: The Case of Gene Therapy Research," *Organization Science* (25:3), pp. 850-867.

Tambe, P., and Hitt, L. M. 2012. "The Productivity of Information Technology Investments: New Evidence from It Labor Data," *Information Systems Research* (23:3-part-1), pp. 599-617.

Webb, M. 2019. "The Impact of Artificial Intelligence on the Labor Market," *Available at SSRN 3482150*.

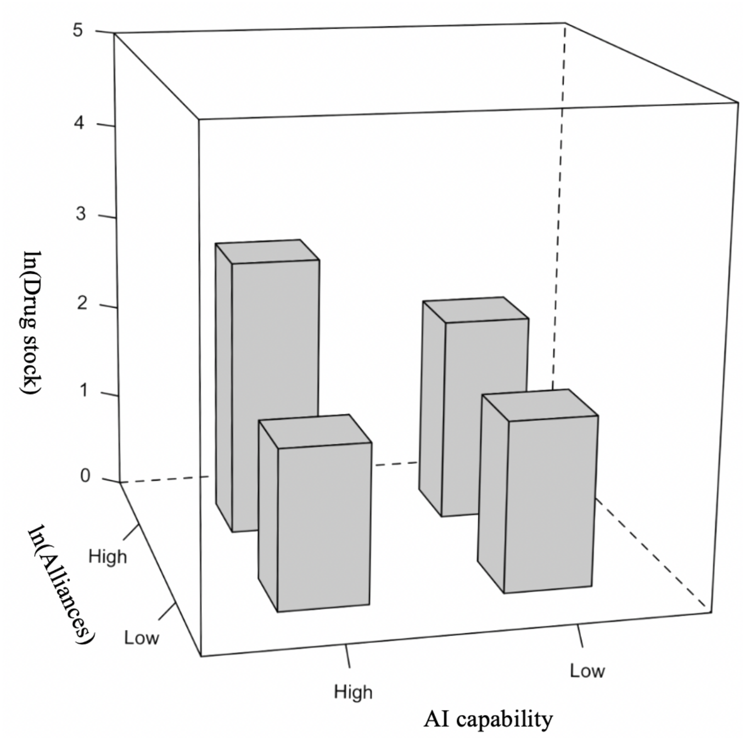
Williamson, O. E. 1991. "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives," *Administrative Science Quarterly*, pp. 269-296.

WIPO. 2019. *WIPO Technology Trends 2019: Artificial Intelligence*, Geneva: World Intellectual Property Organization. (www.wipo.int › edocs › pubdocs › wipo_pub_1055).

Wu, L., Hitt, L., and Lou, B. 2020. "Data Analytics, Innovation, and Firm Productivity," *Management Science* (66:5), pp. 2017-2039.

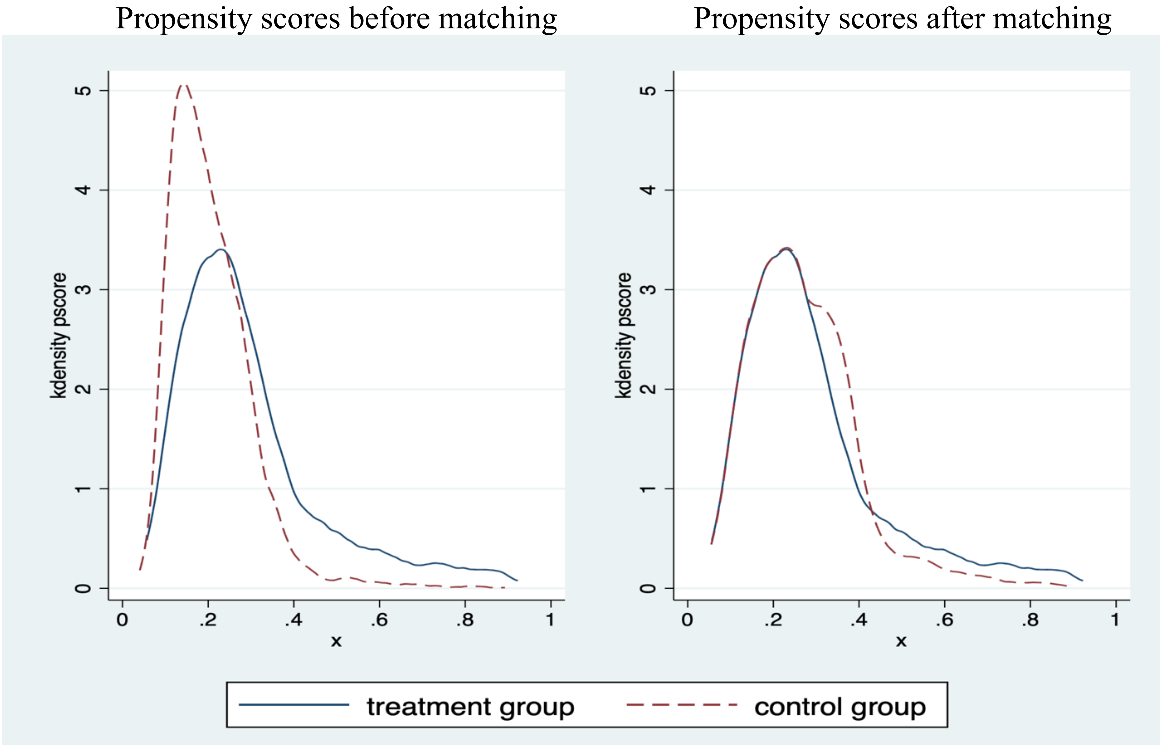
Wu, L., Lou, B., and Hitt, L. 2019. "Data Analytics Supports Decentralized Innovation," *Management Science* (65:10), pp. 4863-4877.

Figure 1: Model-free evidence for an AI resources-R&D alliance complementarity



This figure displays the average logged number of drugs developed by type of firm (e.g., firms with high levels of AI resources and high levels of alliances vs. firms with low levels of AI resources and low levels alliances). The median level of number of alliances is zero. Thus, "Low" values of the number of alliances indicate zero alliances (at the firm level), while "High" values are greater-than-zero alliances. The dichotomy of AI resources (i.e., low vs. high) are also defined by a median split.

Figure 2: Propensity score predicting an alliance before and after matching



The matched sample was created using 1:1 nearest neighbor matching on the probability of forming an alliance at time *t* ("the propensity score") for firms with an alliance (the "treatment" group), compared to firms that did not have an alliance at time *t* (the "control" group).

Table 1. Summary statistics and correlation table (n = 4,778)

Variable	Mean	Std dev.	Min	Max	1	2	3	4	5	6	7	8
					Pairwise correlations							
1. Drugs developed	15.33	37.35	0	561								
2. AI IP	1.22	7.31	0	142.70	0.21							
3. AI skills	16.12	143	0	4,269	0.44	0.39						
4. Alliances	0.61	2.21	0	42	0.76	0.20	0.45					
5. Patents	288.50	1,202	0	21,145	0.62	0.55	0.43	0.55				
6. Public company	0.42	0.49	0	1	0.19	0.11	0.11	0.12	0.16			
7. Firm age	23.81	28.33	0	347	0.23	0.16	0.17	0.23	0.30	0.06		
8. Employees	4,997	15,951	1	316,320	0.51	0.28	0.32	0.41	0.52	0.11	0.21	
9. R&D spend ($B)	3.56	12.10	<0.01	289	0.10	0.05	0.04	0.08	0.07	0.08	0.03	0.08

Notes: *AI resources* is standardized to be a mean 0, standard deviation 1, variable. For ease of interpretation, summary statistics are reported before taking logs.

Table 2. AI, alliances, and innovation: Main results

DV =Drugs developed

	(1)	(2)	(3)	(4)	(5)	(6)
Alliances	7.62*	6.37*	7.15*	5.72*	7.10*	6.73*
	(1.71)	(1.77)	(1.68)	(1.70)	(1.66)	(1.63)
AI IP	11.60*	10.20*				
	(5.03)	(4.89)				
AI IP x Alliances		***4.56****				
		(1.99)				
AI skills			4.26*	2.89+		
			(1.35)	(1.48)		
AI skills x Alliances				***3.27****		
				(1.55)		
AI resources					8.54*	6.77*
					(2.29)	(2.24)
AI resources x Alliances						***4.17****
						(1.75)
Patents	12.10*	12.10*	12.50*	12.40*	12.00*	11.90*
	(2.44)	(2.44)	(2.41)	(2.40)	(2.41)	(2.41)
Public company	13.20*	13.20*	12.60*	12.40*	12.40*	12.30*
	(4.86)	(4.86)	(4.84)	(4.84)	(4.84)	(4.84)
Firm age	79.30*	79.60*	79.20*	79.40*	79.20*	79.60*
	(9.51)	(9.51)	(9.50)	(9.47)	(9.45)	(9.43)
Employees	-0.49	-0.47	-0.52	-0.54	-0.54	-0.54
	(0.56)	(0.56)	(0.56)	(0.56)	(0.56)	(0.56)
R&D spend	0.83	0.82	0.73	0.72	0.76	0.73
	(0.54)	(0.54)	(0.54)	(0.54)	(0.54)	(0.54)
Year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
n	4,778	4,778	4,778	4,778	4,778	4,778
R^2	0.94	0.94	0.94	0.94	0.94	0.94

All coefficient and standard errors are multiplied by 100 for presentation purposes. All variables enter in logs, except *AI resources*, *Public status* and the fixed effects. Robust standard errors are clustered by firms. * $p<0.05$, + $p<0.1$

Table 3. Comparison of firm characteristics before and after matching

	Before Matching			After Matching		
	(1)	(2)		(3)	(4)	
Variable	Firms with an alliance	Firms without an alliance	P-value (1) - (2)	Firms with an alliance	Firms without an alliance	P-value (3) - (4)
AI IP	0.38	0.19	<0.01	0.38	0.33	0.16
AI skill	0.71	0.19	<0.01	0.71	0.43	<0.01
Patents	4.38	3.63	<0.01	4.38	4.20	0.05
Public company	0.52	0.39	<0.01	0.52	0.52	0.85
Firm age	3.07	2.81	<0.01	3.07	3.00	0.08
Employees	6.56	5.88	<0.01	6.56	6.46	0.41
R&D spend	20.20	19.88	<0.01	20.20	20.20	0.96
Year: 2011	0.08	0.11	<0.01	0.08	0.07	0.59
Year: 2012	0.07	0.12	<0.01	0.07	0.08	0.43
Year: 2013	0.10	0.12	0.03	0.10	0.09	0.69
Year: 2014	0.15	0.11	<0.01	0.15	0.15	0.90
Year: 2015	0.16	0.11	<0.01	0.16	0.17	0.71
Year: 2016	0.16	0.11	<0.01	0.16	0.16	0.95
Year: 2017	0.13	0.10	0.05	0.13	0.13	0.89
Year: 2018	0.08	0.12	<0.01	0.08	0.08	0.55
Year: 2019	0.08	0.10	0.06	0.08	0.08	0.86
n	1,050	3,728		939	939	

The matched sample was created using 1:1 nearest neighbor matching on the probability of forming an alliance at time *t* ("the propensity score") for firms with an alliance (the "treatment" group), compared to firms that did not have an alliance at time *t* (the "control" group). Year: 2010 is dropped because explanatory variables are all lagged by one for matching. The F-test on the joint significance of the differences in the means of the covariates is strongly statistically significant before matching, but not statistically significant at conventional levels (p = 0.17) after the matching.

Table 4. AI, alliances, and innovation: 2SLS

	(1-I)	(1-II)	(2)
Dependent variable	AI resources	AI resources x Alliances	Drugs developed
Sample	Matched	Matched	Matched
Specification	2SLS 1st stage	2SLS 1st stage	2SLS 2nd stage
AI resources x Alliances			***7.51*** $^{+}$
			(4.17)
# of neighboring firms w/AI resources	0.39 *	-0.22	
	(0.16)	(0.17)	
# of neighboring firms w/AI resources x Alliances	-0.08	0.38 $^{+}$	
	(0.06)	(0.22)	
Controls	Y	Y	Y
Year FE	Y	Y	Y
Firm FE	Y	Y	Y
n	1,878	1,878	1,878
R^2	0.94	0.94	n/a

All coefficient and standard errors are multiplied by 100 for presentation purposes. Controls include all the variables from Table 2 column 6. All variables enter in logs, except *AI resources*, *Public status*, the instrument (number of neighboring firms with AI resources) and the fixed effects. Column 1 (I-II) and Column 2 use the matched sample from propensity score matching. The associated F-statistic for the first stage is 27.85. Robust standard errors are clustered by firms. * $p<0.05$, $^{+}$ $p<0.1$

Table 5. AI resources and R&D alliances

	(1)	(2)	(3)	(4)	(5)	(6)
DV	Alliance in year *t*	Alliance in year *t*	Alliances	Alliance in year *t*	Alliance in year *t*	Alliances
Specification	OLS	Dif-in-dif	Dif-in-dif	2SLS	2SLS	2SLS
AI resources	***6.17******	***5.91******	***6.68******	***8.85******	***24.80****+*	***32.60****+*
	(0.66)	***(2.22)***	***(2.35)***	***(1.60)***	***(13.60)***	***(18.60)***
Patents	1.71*	-0.70	-0.60	1.26*	-2.44	-3.00
	(0.36)	(1.27)	(1.32)	(0.44)	(1.78)	(2.10)
Public company	5.94*	0.95	2.12	5.31*	-1.36	-0.95
	(1.22)	(2.90)	(2.93)	(1.26)	(3.08)	(3.43)
Firm age	2.93*	-4.56	-5.94	2.65*	-4.66	-6.08
	(0.81)	(4.86)	(4.96)	(0.83)	(4.58)	(4.80)
Employees	0.86*	0.03	0.20	0.69*	-0.14	-0.05
	(0.25)	(0.37)	(0.37)	(0.26)	(0.38)	(0.41)
R&D spend	0.14	-0.23	-0.19	0.05	-0.31	-0.29
	(0.22)	(0.38)	(0.36)	(0.22)	(0.37)	(0.36)
n	4,778	4,778	4,778	4,778	4,778	4,778
R^2	0.09	0.34	0.61	n/a	n/a	n/a
Year FE	Y	Y	Y	Y	Y	Y
Firm FE	N	Y	Y	N	Y	Y

All coefficient and standard errors are multiplied by 100 for presentation purposes. All variables enter in logs, except *AI resources*, *Public status*, the fixed effects and the dependent variable in columns 1, 2, 4 and 5. The instrument was strong in the first stage of the 2SLS results (available upon request). Robust standard errors are clustered by firms in specifications with firm fixed effects. * $p<0.05$, $^{+}$ $p<0.1$

Table 6. AI and knowledge flows within alliances

	(1)	(2)	(3)	(4)
DV	Knowledge flows	Knowledge flows	Knowledge flows	Knowledge flows
Specification	Dif-in-Dif	Dif-in-dif	Dif-in-dif	2SLS
AI IP	***10.40**** ***(4.77)***			
AI skills		***3.26**** ***(1.51)***		
AI resources			***8.07**** ***(2.94)***	***18.00**** ***(8.54)***
Patents	1.11 (1.10)	2.18* (1.06)	1.27 (0.97)	-0.92 (1.23)
Public company	-3.71 (9.82)	-4.22 (9.21)	-5.06 (9.62)	-4.86 (3.52)
Firm age	-6.88 (10.10)	-8.19 (10.40)	-7.47 (10.00)	0.67 (3.54)
Employees	-0.49 (0.30)	-0.54 (0.36)	-0.46 (0.32)	0.21 (0.28)
R&D spend	0.04 (0.28)	-0.22 (0.32)	-0.32 (0.34)	-0.65 (0.46)
Year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
n	2,755	2,755	2,755	2,755
R^2	0.34	0.33	0.34	n/a

All coefficient and standard errors are multiplied by 100 for presentation purposes. The sample includes all alliances in the data used in the previous tables. All RHS variables enter in logs, except *AI resources*, *Public status* and the fixed effects. Knowledge flows are measured as the natural logarithm of the difference in firms' citations to their partners before and after alliances. The instrument was strong in the first stage of the 2SLS estimator (available upon request). Robust standard errors are clustered by firms. * $p<0.05$

Appendix: Keywords used to identify AI-related patents or skills

3d imaging	conditional random field	genetic algorithm	logic program	motion capture	robot
adaboost	data mining	graphical model	logic system	natural language	scikit-learn
anomaly detection	decision tree	hadoop	logistic regression	neural network	semantic analysis
artificial intelligence	deep learning	hidden markov	machine intelligence	objective function	stochastic gradient descent
augmented reality	defuzzification	hyperspectral imaging	machine learning	pattern recognition	supervised learning
cloud computing	dimensionality reduction	ibm watson	maximum entropy	predictive analysis	support vector machine
cluster analysis	evolvable hardware	inference engine	maximum a posteriori	predictive model	tensorflow
computational biology	expert system	information extraction	maximum likelihood	pytorch	unsupervised learning
computational control	feature selection	keras	mechatronic	random forest	virtual reality
computer vision	fuzzy logic	knowledge base	motif discovery	reinforcement learning	xgboost